JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2009, **91,** 127–141 NUMBER 1 (JANUARY)

ACQUIRED EQUIVALENCE CHANGES STIMULUS REPRESENTATIONS

M. MEETER[1], D. SHOHAMY[2], AND C.E. MYERS[3]

[1]DEPT. OF COGNITIVE PSYCHOLOGY, VRIJE UNIVERSITEIT AMSTERDAM
[2]DEPT. OF PSYCHOLOGY, COLUMBIA UNIVERSITY
[3]DEPT. OF PSYCHOLOGY, RUTGERS UNIVERSITY

Acquired equivalence is a paradigm in which generalization is increased between two superficially dissimilar stimuli (or antecedents) that have previously been associated with similar outcomes (or consequents). Several possible mechanisms have been proposed, including changes in stimulus representations, either in the form of added associations or a change of feature salience. A different way of conceptualizing acquired equivalence is in terms of strategic inference: Confronted with a choice on which it has no evidence, the organism may infer from its history of reinforcement what the best option is, and that inference is observed as acquired equivalence. To test this account, we combined an incremental learning task with an episodic memory test. Drawings of faces were made equivalent through acquired equivalence training, and then paired with words in a list learning paradigm. When participants were asked to recognize specific face-word pairings, they confused faces more often when they had been made equivalent. This suggests that prior acquired equivalence training does influence how memories are coded. We also tested whether this change in coding reflected acquisition of new associations, as suggested by the associative mediation account, or whether stimuli become more similar through a reweighting of stimulus features, as assumed by some categorization theories. Results supported the associative mediation view. We discuss similarities between this view and exemplar theories of categorization performance.

Key words: memory, learning, hippocampus, conditioning, acquired equivalence, humans

In an acquired equivalence task, an organism learns that two or more stimuli are equivalent in terms of being mapped onto the same outcomes or responses. This is referred to as *functional equivalence*, as the grouping of stimuli is not based on stimulus characteristics, but only on a functional characteristic such as predicting the same outcome. In a typical experiment, two antecedent stimuli, A1 and A2, are first both followed by a reward, while another antecedent stimulus, B, is not. When, in a second transfer stage, A1 is paired with a shock, the conditioned fear will generalize from A1 to A2, and not to B (Ward-Robinson & Hall, 1999, Expt. 2). It is as if A1 and A2 have become equivalent to the animal because they predicted the same outcome (or consequent) in the first stage of the experiment; subsequent learning about A1 then transfers easily to A2. Functional equivalence is interesting in its own right, but it has also been studied as a way in which humans and nonhuman animals learn to categorize (Urcuioli, 2001), and has been suggested to underlie the learning of symbolic reference (Sidman, 1994; Tonneau, 2001).

Although often studied in animal learning paradigms (e.g., Bonardi, Rey, Richmond, & Hall, 1993; Honey & Hall, 1989, 1991), functional equivalence can also be found in humans using, for example, the 'Fish' task (Myers et al., 2003). On each trial of the 'Fish' task, participants see a face and a pair of fish, and have to learn through trial and error which of the fish goes with that face (Figure 1). There are four faces (A1, A2, B1, B2), referred to as antecedents, and four possible fish (X1, X2, Y1, Y2), referred to as consequents (more common terminology in the behavior-analytic literature is 'sample stimuli' and 'comparison stimuli', respectively). In the initial training stages participants learn that, given face A1 or A2, the correct answer is to choose fish X1 over fish Y1; given face B1 or B2, the correct answer is to choose fish Y1 over fish X1. During this phase, participants learn that face A1 and A2 are equivalent with respect to the associated fish; face B1 and B2 are likewise equivalent. Next, participants learn a new set of pairs: Given face A1, choose fish X2

This research was supported by a VENI grant to the first author from the Netherlands Society for Scientific Research (NWO).

Correspondence should be addressed to Martijn Meeter, Dept. of Cognitive Psychology, Vrije Universiteit Amsterdam, Vd Boechorststraat 1, 1018 BT Amsterdam, The Netherlands (e-mail: m@meeter.nl).

doi: 10.1901/jeab.2009.91-127

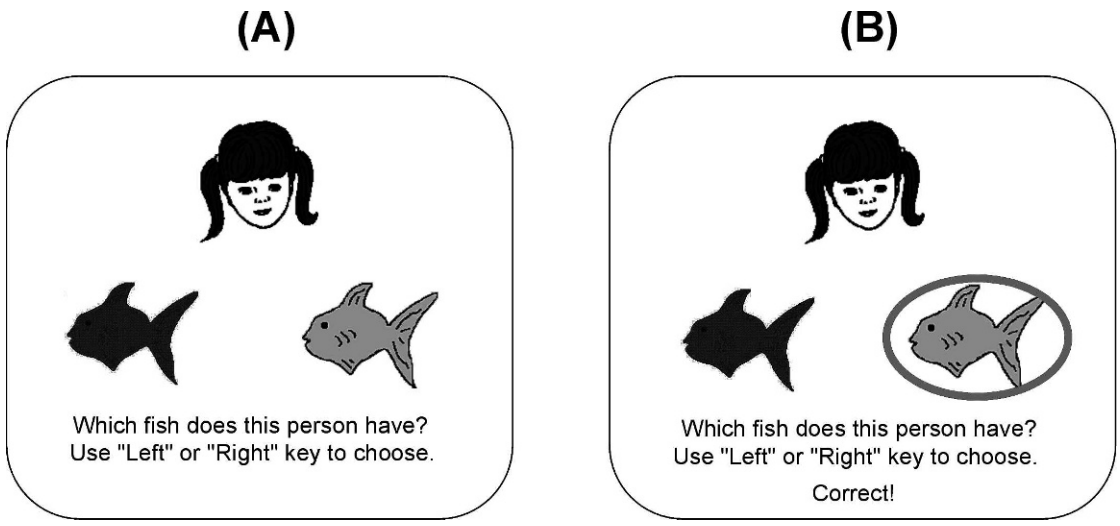

Fig. 1. Example screen events during one trial. (A) Stimuli appear. (B) Participant responds and corrective feedback is given.

over Y2, and given face B1, choose fish Y2 over X2. Participants are then given a transfer test: Given face A2 or B2, will subjects choose fish X2 or Y2? Having learned that faces A1 and A2 are equivalent, participants may generalize from learning that A1 goes with Y1 to that A2 also goes with X2; the same holds for B2 (equivalent to B1) and Y2 (associated with B1). Healthy adults (Hall, Mitchell, Graham, & Lavis, 2003; Myers et al., 2003), children (Goyos, 2000; Schenk, 1994) and retarded individuals (Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989) all reliably make this type of generalization. Similar behavior can be observed in nonhuman species including rats and pigeons (e.g., Bonardi et al., 1993; Honey & Hall, 1991; Urcuioli, 2001).

In the behavior-analytic literature, functional equivalence is interpreted as equivalence training grouping stimuli into equivalence classes based on common reinforced choices (e.g., Sidman & Tailby, 1982; Tonneau, 2001). By virtue of pairing with outcomes, antecedents A1 and A2 are altered to form one *equivalence class*, while B1 and B2 form another. When confronted with a new, untrained choice in the transfer test, training with other class members may generalize to this new choice. For example, when two stimuli, A1 and A2, are paired with the same outcome (e.g., X1), and A1 is subsequently paired with another outcome (X2), this training will generalize to the other member of the equivalence class. Generalization between stimuli within an equivalence class is enhanced, while generalization between stimuli in different equivalence classes is reduced. The underlying process is often referred to as *transfer of function* from A1 to A2. Equivalence classes can come about not only by common reinforced choices, but also when members of a class share a common reinforcer (e.g. Dube et al., 1989). Antecedents and consequents also readily form equivalence classes (Markham, Dougher, & Augustson, 2002), as may stimuli and the responses that must be given to them (though see Urcuioli, Lionello-DeNolf, Michalek, & Vasconcelos, 2006, for contrary evidence with pigeons).

At a cognitive level, there are two ways to understand equivalence classes. On the one hand, generalization within them may reflect a *strategic inference*. Since organisms do not have any evidence to base their choice on in transfer tests, they may assume that it is most advantageous to base their choice on training with other class members. In the example given above, they may infer that, in the absence of better evidence, A2 is more likely than not to require a choice of X2, given that A1 requires a choice of X2.

Alternatively, members of an equivalence class may be stored in memory in a way that increases generalization between them. In-

deed, many explanations for acquired equivalence are based on the idea that, during initial training, the representations of stimuli paired with the same outcome or consequent are modified to increase subsequent generalization between them (e.g., Gluck & Myers, 2001; Hall et al., 2003; Myers & Gluck, 1996). Conversely, representations of stimuli that are associated with different outcomes should become less alike.

Two mechanisms have been proposed to underlie such *representational change*. One hypothesis is that representations of equivalent stimuli become more similar because overlapping features become more salient. In the fish task, faces have three salient features (age, gender, hair color), of which equivalent faces share one (the specific shared feature is counterbalanced over participants). For example, suppose that two faces with yellow hair are paired with the same consequent, while two brown-haired faces, paired with a different consequent, form a second equivalence class. Equivalence training might make hair color more salient, emphasizing the differences between equivalence classes, while deemphasizing other features (gender and age) that vary within an equivalence class. This would result in antecedents within one equivalence class being perceived as more similar to each other, and less similar to antecedents from other equivalence classes. Hypotheses of this kind, which emphasize selective attention to one or more features that control responding, have been proposed by theories of animal learning (Mackintosh, 1975), have received experimental support (e.g., Foree & Lo-Lordo, 1970; Mackintosh, 1965; Reynolds, 1961), and are common currency in the field of categorization (Kruschke, 1979, 1992; Medin & Schaffer, 1978). Indeed, acquired equivalence tasks are similar to categorization tasks in that stimuli are grouped into classes according to the outcome or response they evoke. In many acquired equivalence studies with human participants, the first stage of training also uses verbal labels, making it equivalent to a categorization task (e.g., participants may be asked to discover that one verbal label applies to A1 and A2, and another to B1 and B2).

This hypothesis, which we will call the *feature salience account* of acquired equivalence, is intuitively appealing. If, for example, hair color can be used to remember which choice to make in the presence of which face, it seems natural to suggest that participants use hair color to code the faces when remembering the choices to be made.

An alternative hypothesis was presented by Hall and colleagues (Hall et al., 2003; Ward-Robinson & Hall, 1999), building on earlier ideas by Hull (1939). They propose that acquired equivalence training adds an association to stimulus representations. Retrieval of this association later in training then leads to the transfer. In the Fish paradigm, participants might first learn to associate faces A1 and A2 with fish X1. In a later phase, participants are trained to pair face A1 with fish X2. During that training, the previously learned association of face A1 and fish X1 will be activated, leading fish X1 and X2 to become associated, by virtue of their pairing with the same consequent. When face A2 is now shown, it will activate its prior association with X1, which in turn activates X2; thus, the participant will choose X2 in the presence of A2.

Hall et al. (2003) found evidence for this *associative mediation* account of acquired equivalence. In one experiment, they linked stimuli A1 and A2 to color patches (e.g., a square in a distinct red color). They found that when A1 was coupled with a response in a second training stage, this generalized not only to A2 but also to the color patch. This is what would be expected by the associative mediation account, namely: Generalization can occur between consequents that had previously been paired with the same antecedent. However, the same research group also found evidence for the feature salience account. In an acquired equivalence task, participants learned about four antecedents: two snowflake patterns and two color patches. These were paired with consequents so as to form equivalence classes that either shared a clear feature (two snowflake patterns in one class, two color patches in the other) or did not (one snowflake and color patch in each class). As would be predicted by the feature salience account, the equivalence training had a much stronger effect on later generalization when the equivalent antecedents shared a clear feature than when they did not (Bonardi, Graham, & Hall, 2005). The experiment was set up in such a way that associative mediation was excluded as an explanation for these results.

In the studies reported here, we first sought to clarify whether acquired equivalence training would lead to responses that could not be based on strategic inferences (Experiment 1). If representations are indeed being changed during training, this should also show when they are probed in different ways, such as in an explicit memory test. In particular, if representations of two stimuli have become more similar through equivalence training, then they should also become more confusable in episodic memory. We tested this by combining equivalence training with an episodic associative recognition task. If acquired equivalence changes representations, participants should make more errors on items trained for equivalence (critical lures) than on items that had not been trained for equivalence (control lures). If, on the other hand, acquired equivalence relies on inferences and not on representational change, then performance on the episodic memory test should not be impacted by acquired equivalence. This is so because in the associative recognition task, a response based on the equivalence of stimuli is an error. Since participants have no incentive to make an error, they have no incentive to infer responses from the equivalence training. This contrasts with traditional tests of transfer of function, in which responses based on equivalence training are never predefined as errors; either no information is given about which answers are correct, or participants receive feedback only after giving the response. In such contexts, responding on the basis of experience with other members of an equivalence class could be argued to be a rational strategy.

Next (Experiment 2), we reversed the order of equivalence training and episodic memory task, to determine whether the effects of acquired equivalence resided at training or testing. To preview results, Experiment 1 showed that acquired equivalence training could indeed influence performance in an episodic memory test, while Experiment 2 showed that the effects occur during training, not testing.

Having obtained support for a representational account of acquired equivalence, we wanted to contrast the two possible mechanisms for representational change (feature salience vs. associative mediation) in a design in which the two accounts would lead to opposite predictions (Experiment 3). Results were consistent with the associative mediation account, rather than the feature salience account.

EXPERIMENT 1

Experiment 1 tested whether equivalence training altered stimulus representations, by combining equivalence training with an episodic memory task (see Figure 2). Participants underwent acquired equivalence training on fish–face pairs, and were then given a generalization test. They then studied words in the presence of the previously trained face antecedents. Participants were only instructed to study the words, but were subsequently tested on their recognition of the word–face pairings (associative recognition). Finally, the face–fish generalization test was repeated to document that face–word training had not disrupted the equivalencies learned earlier.

During the recognition test, participants were confronted with studied word–face pairs as well as with two kinds of lures: *critical lures* in which a word that had been studied with one face was now shown with an equivalent face (e.g., a word that had been paired with face A1 at training was shown at test with equivalent face A2), and *control lures* in which the face at test was not equivalent to the face at study (e.g., a word combined with face A1 at training was shown at test with nonequivalent face B1). If acquired equivalence changes representations that are also involved in episodic memory, then participants should make more errors on critical lures than on control lures. If, on the other hand, acquired equivalence relies on inferences and not on representational change, no such difference should be found.

Method

Subjects

Participants were 34 students at Rutgers University, who received class credit in an introductory psychology class in exchange for their participation. Twenty participants were female and 14 male; mean age was 19 years (range 17–27). For 2 participants, data from the second transfer test were lost due to computer failure; data from all previous stages were analyzed for these participants. All

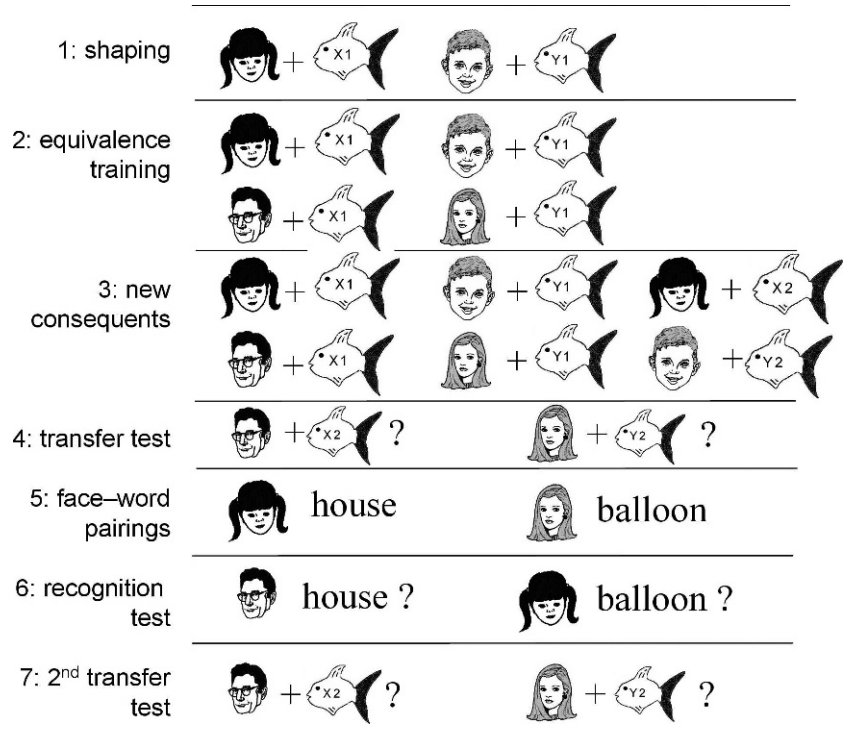

Fig. 2. Stages of Experiment 1. In the first three stages, participants learn via trial and error which faces ''have'' which fish (shown is one set of face–fish pairings, but these were different for each participant). In transfer tests (stages 4–7), acquired equivalence is tested by letting participants decide on unseen face–fish pairings without receiving feedback. In the face–word study, faces are paired with words. Memory for these pairings is subsequently tested in an associate recognition test. In this test, the words can either be shown with the face that they were paired with in the study stage, with a face that is equivalent to the on it was paired with (left pair, a *critical lure*), or with a face that was not made equivalent (right pair, a *control lure*).

participants signed statements of informed consent before the initiation of any behavioral testing, and research procedures conformed to the regulations established by the Federal Government and by Rutgers University.

Apparatus and Stimuli

Behavioral testing was automated on a computer. The participant was seated at a comfortable viewing distance from the screen (approximately 700 mm). The keyboard was masked except for two keys, labeled ''LEFT'' and ''RIGHT'', which the participant could press to record a response.

Four drawings of faces (man, woman, girl, boy) served as the antecedent stimuli. The boy and woman had yellow hair while the girl and man had brown hair. Thus, each antecedent had three obvious, binary-valued features: age (adult vs. child), gender (male vs. female) and hair color (yellow vs. brown); each antecedent shared exactly one feature with each other antecedent. For each participant, the four face drawings were randomly assigned to be antecedents A1, A2, B1, and B2. For each participant, drawings of a fish colored red, blue, green, and yellow were randomly assigned to X1, X2, Y1, and Y2. Faces and fish appeared about 3 cm high on the computer screen.

For face–word pairing, the 16 words used were: pencil, farmer, curtain, house, parent, garden, turkey, mountain, river, bell, coffee, nose, hat, school, moon, drum. Four words were paired with each antecedent, with face–word assignment randomized across subjects. Words were presented on the screen in black lowercase font (Arial, 24-point), about 1 cm high.

Procedure

At the start of the experiment, on-screen instructions stated: ''You will see drawings of

people who each have some pet fish. Different people have different kinds of fish. Your job is to learn which kinds of fish each person has. At first, you will have to guess.'' The experimenter read these instructions aloud to the participant.

In all, the experiment consisted of seven stages (Figure 2). Stages 1–4 were the same as previously described in Myers et al. (2003). The first three stages were face–fish training. On each trial, the screen showed a single antecedent (face), two consequents (fish) presented side-by-side in randomized left–right order, and a prompt: ''Which fish does this person have? Use the LEFT or RIGHT key to choose.'' An example of screen appearance at the start of a trial is shown in Figure 1. The participant responded by pressing one of the two labeled keys. The selected consequent (fish) was circled, and corrective feedback was given. The stimuli remained on the screen until the participant responded; feedback was shown for 1 s, and there was a 1-s intertrial interval, during which the screen was blank.

There were three stages of training, each with increasing numbers of trial types as shown in Figure 2. In the first shaping stage, trials presented one of two faces (A1 or B1) and a pair of fish (X1 and Y1), and the participant learned to pair the correct fish with each face (A1 with X1 and B1 with Y1). Since the fish could appear in either left–right order, this made four trial types in Stage 1 (A1 with X1, Y1; A1 with Y1, X1; B1 with X1,Y1; B1 with Y1,X1). A block of trials in Stage 1 consisted of one of each of these four trial types, in pseudorandom order. Shaping continued for a maximum of eight blocks (32 trials) or terminated early if the participant made eight consecutive correct responses.

The second equivalence training stage then began, without warning to the participant. In equivalence training, the previously learned trial types were intermixed with trials on which one of two new faces (A2 or B2) was presented with the familiar fish (X1 and Y1). Since, again, the fish could appear in either left–right order, there were eight trial types, four old and four new. Each block consisted of eight trials, one with each trial type, in pseudorandom order. Equivalence training continued until the participant reached a criterion of eight consecutive correct responses, or for a maximum of eight blocks (64 trials). The third ''new consequents'' stage then began, without warning to the participant. Here, on each trial, a familiar face (A1 or B1) was presented with two new fish (X2 and Y2). Again, the fish could appear in either left–right ordering, and these new trial types were interleaved with the previously trained ones; thus, there were 12 trial types in each block. Training with new consequents continued until the participant made 12 consecutive correct responses or until a maximum of eight blocks (96 trials) had been completed.

At the conclusion of training, the following instructions appeared: ''Good! In this part of the experiment, you will need to remember what you have learned so far. You will NOT be shown the correct answers. At the end of the experiment, the computer will tell you how many you got right. Good luck!'' A transfer test (stage 4) followed, which consisted of 16 trials: all 6 trial types from the acquisition stages plus the 2 new test trial types (A2 paired with X2 or Y2, and B2 paired with X2 or Y2), with the consequents in each possible left–right ordering. On each trial, the screen showed one face and two fish; the fish chosen by the participant was circled, but no corrective feedback was given. Trial order was random for each participant.

In the fifth face–word study stage, participants were instructed to study words. On each trial, a single face (A1, B1, A2, or B2) appeared in its usual position, with a single word underneath. Each pairing was shown for 5 s. Since each face was paired with four words total, there were 16 trial types. Face–word pairings included two passes through a block of 16 trial types, with order randomized within a block, for a total of 32 trials. Participants were instructed to study the words for a later ''memory test'', and were not given instructions on how to process the faces.

The sixth stage was an associative recognition test, containing 16 trials—one with each of the 16 studied words. Each word appeared with one of the four face stimuli, and participants were asked whether that particular face–word pairing had occurred during study. On eight of the trials, the word–face pair had indeed been presented during training; errors here were *misses*. On the other eight trials, a word appeared together with a face with which it had not previously been paired; errors here were *false positives*. On four of these, the word

Table 1

Results from the two transfer tests in Experiment 1 (stage 4 and stage 7) and the one transfer test in Experiment 2.

	Expt 1		
	1st test	2nd test	Expt 2
old pairs	0.02	0.03	0.03
transfer pairs	0.15	0.08	0.16

Note. ''Old pairs'' refers to the face–fish pairings trained in stages 3–4. ''Transfer pairs'' refers to the pairings in which a face and a fish are paired that did not occur in the stages 3–4. Shown is the proportion of errors on the 16 trials of the transfer tests.

appeared paired with a face that had been made equivalent with the face presented during study; these were *critical lures*. On the other four trials, words were paired with nonequivalent faces; these pairs were *control lures*.

The final stage was a second transfer test, identical to the first one. It was intended as a posttest to confirm that learning the word–face pairs had not disrupted the original face–fish associations.

Results

All participants learned all trained face–fish pairs within the maximum allowable number of trials. Table 1 shows the results from the transfer tests. To make errors amenable to standard parametric tests, we transformed error proportions to a form suitable for analyses of variance[1] (Winer, Brown, & Michels, 1991). In both transfer tests, very few errors were made on the trained pairs. There were more errors on the transfer pairs than on the old pairs in the first transfer test (two-tailed paired-samples t-test, $t(34) = 2.78$, $p < .01$), but many fewer than the 50% that would be expected by chance, $t(34) = 6.34$, $p <. 001$ one-tailed. On the final transfer test, there were again many fewer errors than could be expected by chance, $t(31) = 11.3$, $p < .001$ one-tailed. There was no difference between old and new pairs, $t < 1$, due to slightly fewer errors on the new pairs in the final transfer test and slightly more errors on the old pairs, although neither difference was significant (first vs. final test for old pairs: $t(31) = 1.01$, $p = .32$; for new pairs: $t(31) = 1.6$, $p = .12$, both two-tailed). Both transfer tests show a standard acquired equivalence effect, while the second test confirms that the learned equivalencies were maintained during the episodic learning and test stages.

Figure 3 shows the results from the recognition memory test (sixth stage). Consistent

[1] $x = 2 \arcsin \sqrt{(p + 1/2n)}$, where x is the transformed variable, p is the proportion, and n is the number of observations underlying the proportion.

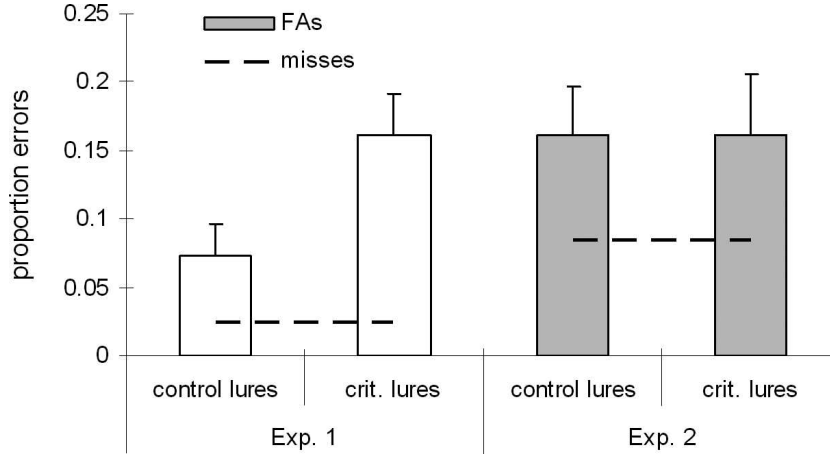

Fig. 3. Error rates in the episodic memory test in Experiment 1 and Experiment 2. Participants had to judge whether words and faces had been shown together at study. Dashed bars give the number of intact pairs erroneously labeled as new (''misses''). ''Control lures'' are word–face pairings in which the face is nonequivalent to the face shown with the word at study. ''Critical lures'' are word–face pairings in which the face was made equivalent during equivalence training to the face shown with the word at study. Error bars give 1 SEM.

with our hypothesis, participants made more errors on critical lures than on control lures, $t(33) = 2.05$, $p = .024$ one-tailed). Eleven subjects made only critical lure errors, while just 2 made only control lure errors and 6 made both kinds of errors.

Discussion

The results of Experiment 1 show that acquired equivalence may alter performance on an episodic memory test. Our explanation for this is that equivalence training alters the representations of stimuli, which in turn increases generalization between these stimuli when they are subsequently invoked during face–word training. These representational changes should increase similarity between stimuli within an equivalence class, while decreasing similarity between stimuli in different equivalence classes. These representational differences may then lead to the increased confusability we found for equivalent stimuli as compared to nonequivalent stimuli.

The results cannot be explained in terms of strategic inferences. The participants did not have a strategic incentive to respond on the basis of equivalence classes, as such responses were errors in the task. Even if they had guessed that they would be tested on word–face pairings, they would have had an incentive to code what makes the faces distinctive, not what makes them similar to one another. The results thus seem to support an explanation of acquired equivalence in terms of changed representations.

EXPERIMENT 2

The results of Experiment 1 suggest that equivalence training alters the representations of antecedent stimuli paired with the same consequent, and that these altered representations affect subsequent episodic learning about those stimuli. It is possible, however, that the effect resides not at acquisition, but at retrieval. It may be that the acquired equivalence training changes the ways in which the faces are used as cues in the recognition test. To test for this possibility, we performed a second experiment where the word–face training preceded the acquired equivalence training; the episodic memory test still followed the acquired equivalence training. If acquired equivalence training changes the way episodic memories are coded, it should not affect performance on episodic memories acquired before the equivalence training. If, on the other hand, acquired equivalence training changes the way cues are used at test, we should replicate the findings of Experiment 1.

Method

Participants were 31 students at Rutgers University, who, as in Experiment 1, received class credit for their participation; 15 were female, and 16 male; mean age was 19 years (range 18–24). All participants signed statements of informed consent before the initiation of any behavioral testing. All research procedures conformed to the regulations established by the Federal Government and by Rutgers University.

Stimuli and stages of Experiment 2 were the same as in Experiment 1, but the ordering of the stages was changed. The face–word pairings study stage (fifth stage of Experiment 1) now came first, and thus preceded the acquired equivalence training and the transfer test (Stages 1–4). The recognition test followed the transfer test. There was no final transfer test as in Experiment 1.

Results

As in Experiment 1, all participants learned all face–fish pairs within the maximum allowable number of trials. Table 1 shows the results from the transfer test. We again transformed error proportions to make them amenable to standard parametric tests. Very few errors were made on the trained pairs. There were more errors on the transfer pairs than on the old pairs, $t(30) = 2.44$, $p = .02$ two-tailed, but many fewer than the 50% that would be expected by chance, $t(30) = 6.44$, $p < .001$ one-tailed. The magnitude of acquired equivalence was similar to that obtained in Experiment 1.

Figure 3 shows the results from the recognition memory test. There were more errors on both studied words and lures than in Experiment 1, which is unsurprising given the longer interval between study and test. More importantly, participants did not make more errors on critical lures than on other rearranged words, one-tailed, $t < 1$. Just 3 subjects made only critical lure errors, with 6 making only control lure errors and 9 made

both kinds of errors. Thus, the effect of acquired equivalence on episodic memory obtained in Experiment 1 was abolished if the word–face training preceded the equivalence training.

Discussion

Experiment 1 showed that equivalence training could influence episodic memory involving the same stimuli, by increasing confusability among stimuli that had previously been associated with the same consequent. This effect is in a sense a memory distortion, as increased similarity led to more responses that are defined as errors by the task. Other forms of distortion are false memory effects such as those in the Deese-Roediger-McDermott paradigm (Roediger & McDermott, 1995). In this paradigm, a list of words is learned that are all related to one critical lure. At test, participants tend to recall or recognize the lure as seen, even though it was itself not presented. Accounts based on both learning and retrieval have been proposed, but recent evidence suggests that all or part of the effect resides at the retrieval stage: At test, the lure activates memories of list items, and is therefore mistaken for one (Zeelenberg, Boot, & Pecher, in press). Experiment 2 seems to rule out such retrieval-based accounts for our results, as we did not find any effect of equivalence training on recognition performance when the training occurred after encoding but before retrieval. This suggests that equivalence training changes the representations of the equivalent stimuli, which in turn influences how the stimuli are processed during episodic encoding.

A confounding factor in the comparison between Experiments 1 and 2 is that the delay between episodic study and test was greater in Experiment 2 than in Experiment 1, and that participants therefore made more errors in Experiment 2. It is conceivable that this somehow affected the difference between control and critical lures. However, forgetting swiftly results in the confusion of similar stimuli, whereas discriminations between more distinct stimuli are relatively robust against forgetting (e.g., Kraemer, 1984). The difference in performance on the more similar critical lures and the more dissimilar control lures would thus be more likely to grow with time than to shrink.

From a behavioral perspective, there is no reason to expect that reversal of the order of episodic study and equivalence training would result in different performance. Urcuioli (2001) performed a traditional acquired equivalence task in pigeons, but placed the equivalence training after training with new consequents. He found that this reordering did not affect transfer of function in a transfer test. Also, both Dube et al. (1989) and Goyos (2000) found that training subjects to reassign stimuli to a different equivalence class could change how training with new consequents was applied to those reassigned stimuli. Equivalence training can thus affect already stored associations. That it did not in this experiment strengthens our interpretation that the effect of equivalence training on episodic memory performance resides at study, not at test.

EXPERIMENT 3

Experiments 1 and 2 together provide converging evidence that acquired equivalence does involve change in stimulus representations. There are two viable accounts of what form this representational change might take. The feature-salience account suggests that features that differ between equivalence classes are emphasized (made more salient), making representations of equivalent stimuli more similar. The associative-mediation account suggests that the important change occurs in associations between consequents. Experiment 3 was designed to test these two accounts.

Consider the case in which one feature, say hair color, determines which faces are equivalent in the fish task: Yellow-haired faces are associated with X1 and brown-haired faces with X2. What would happen if the hair color of one of the faces were switched? The feature-salience account would predict that participants would reassign this face to a different equivalence class, as hair color is the feature that determines antecedent equivalence. The associative-mediation account would predict the contrary: As long as the face remains recognizable, it should continue to evoke the same response as before.

Another case in which the two accounts make opposing predictions is when a new antecedent is introduced. The feature-salience account would predict that new antecedents

should be automatically assigned to an equivalence class based on their features. For example, if hair color is the salient feature, a new face with yellow hair should elicit the same responses as other yellow-haired faces. In contrast, the associative-mediation account would assign no special meaning to a single feature, such as hair color; instead, the new face should either be randomly associated with a consequent (as participants simply guess to which class it might belong) or else should be assigned to the class containing stimuli with which it shares the most features overall.

In Experiment 3, we simplified the Fish task to only the shaping and equivalence-training stages. In the experimental condition, equivalence classes were based on hair color (so that yellow-haired faces were mapped to consequent X1, while brown-haired faces were mapped to consequent Y1). This was because hair color is the only feature that can be changed without changing the identity of the face. Hair color is not a weak feature, however, in that participants learn the equivalence task as readily if the grouping is based on hair color as when it is based on a different feature. All learned associations were tested in a third stage, followed by testing on a previously trained face with a hair color change, as well as testing with a new face entirely. In the control condition, equivalence classes were based on age (adult vs. child) or gender (male vs. female) but not hair color, followed by the same transfer test of a previously trained face with a hair color change (see Figure 4 for design and predictions).

Both the feature-salience and mediated-associations accounts predict that, in the control condition, the hair color change will be ignored, and the face should be mapped as previously. In the experimental condition, however, predictions differ. According to the feature-salience account, a hair color change should lead the face to be classified with other faces that have the same hair color; according to the associative-mediation account, the face should be identified as it always was, regardless of a single new feature (hair color).

Method

Participants were 40 students at Rutgers University, who received class credit in an introductory psychology class in exchange for their participation. Six participants were male, and 34 were female; mean age was 19 years (range 17–31). All participants signed statements of informed consent before the initiation of any behavioral testing, and research procedures conformed to the regulations established by the Federal Government and by Rutgers University. Participants were randomly and evenly assigned to the experimental and control conditions.

Procedurally, the shaping and equivalence training stages were similar to Experiment 1: On each trial, one of the antecedent faces was presented together with two possible consequent fish stimuli. The ''new consequents'' stage was omitted. The last phase was a no-feedback phase to test learning of face–fish associations.

Four additional drawings were constructed from the usual faces by switching hair color (see Figure 4). There were two conditions in the experiment. In the experimental condition, assignment of the face drawings to equivalence classes during Stages 1 and 2 was based on hair color; i.e., the yellow-haired woman (A1) and yellow-haired boy (A2) were associated with consequent X1, while the brown-haired girl (B1) and brown-haired man (B2) were associated with consequent Y1. The other features (gender and age) varied across and within equivalence classes. These antecedents and consequents were trained in Stages 1 and 2, exactly as in the previous experiments. In the transfer test trained pairs (A1–X1, B1–Y1) were tested, as well as presenting a new face: a brown-haired boy that could be paired with either X1 or Y1.

In the control condition, for half the participants, equivalence class was based on gender: the brown-haired girl (A1) and yellow-haired woman (A2) mapped to X1; the yellow-haired boy (B1) and yellow-haired man (B2) mapped to Y1. For the other half of participants in the control condition, equivalence class was based on age (adults vs. children): yellow-haired woman (A1) and yellow-haired man (A2) mapped to X1; brown-haired boy (B1) and yellow-haired girl (B2) mapped to Y1. Equivalence training was followed by the testing of the trained pairs (A1–X1 and B1–Y1) as well as a new face. For participants trained on gender, this was a brown-haired man who, according to both accounts, should be mapped to Y1 like the previously trained yellow-haired man. For those trained on age,

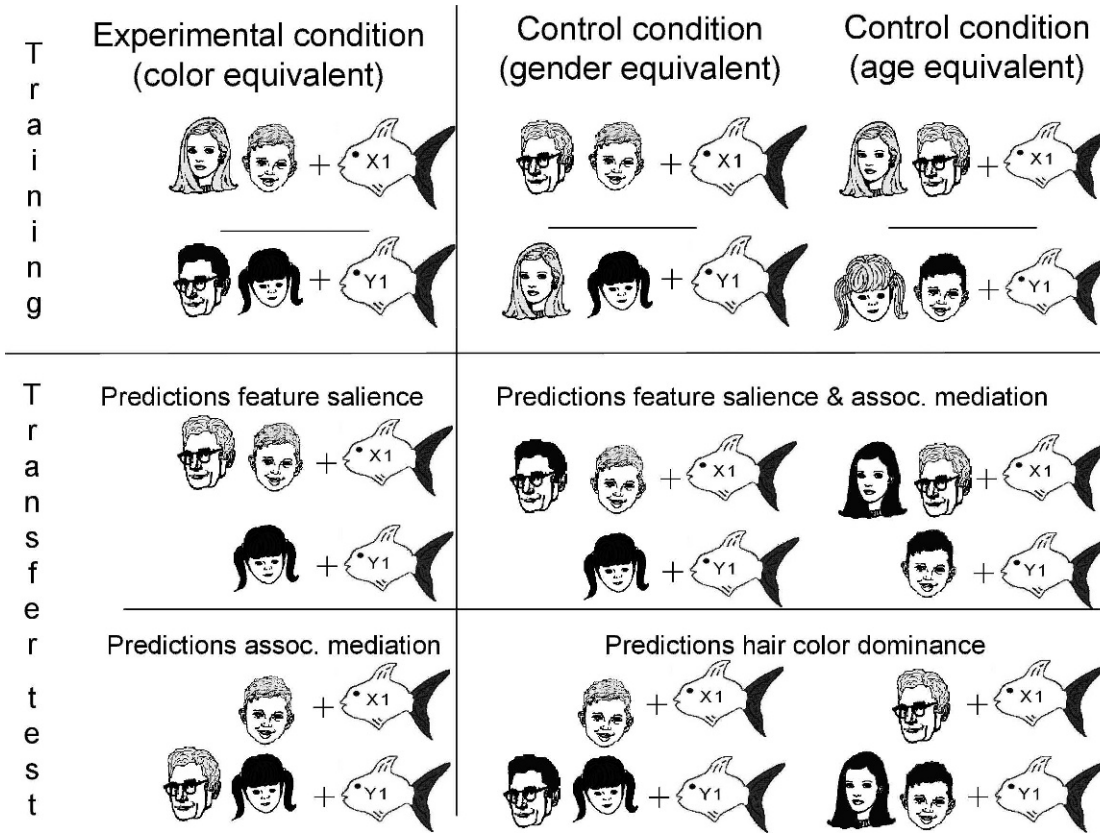

Fig. 4. Design of Experiment 3, and predictions of the two accounts. In the first two stages (Training), participants learn via trial and error which faces ''have'' which fish (shown is one set of face–fish pairings, but these were different for each participant). In the experimental condition, equivalence is based on hair color of the face stimuli, while in the control condition it is based on either gender or age. In stage 3 (Test), participants must decide on face–fish pairings without receiving feedback. Given in the lower part of the figure are the responses as predicted by the feature-salience account, the associative-mediation (assoc.) account, and the idea that hair color is a feature that dominates responding independent of what equivalence training was given.

it was a brown-haired woman who should be mapped to X1, like the other adults. The control condition was designed to test whether hair color was an overwhelmingly salient feature, in which case the new faces should be mapped to the opposite consequents like the only other antecedent with the same hair color.

For all subjects, the transfer test consisted of one presentation of each of the three trial types, with the fish in each possible left–right order, for a total of six trials. Trials were presented in randomized order for each subject.

Results

Within the control condition, no difference between participants in the age and gender subgroups could be found; data from these subgroups were therefore combined. Errors were transformed as in the previous experiments. There was no difference between the conditions on the number of errors in the first or second stage of training, $t(38) < 1$, nor on the errors made on the ''old'' faces in the transfer test [mean number of errors = .15 and .30 for the experimental and control conditions respectively; $t(38) < 1$]. In other words, participants in the control and experimental conditions were equally able to learn and retain equivalencies based on hair color as on age or gender.

Because of the few trials involved, responses to the changed face stimuli were analyzed as follows. Responses could either be based on

Table 2

Number of Experiment 3 participants whose responses to the face with changed hair color were based on either face identity or hair color.

	Experimental condition	Control condition
all responses based on face identity	10	11
mixed responding	4	4
all responses based on hair color	6	5

Note. The feature-salience account predicts hair color responses in the experimental condition, the associative-mediation account predicts identity responses in both conditions.

the identity of the face, or on the hair color. In the first case the response to a face would remain as it was in the training stages, in the latter it changes into what is appropriate for faces with its new hair color. Responses appropriate for faces with the new hair color will be called ''hair color responses'', and were the central dependent variable. The feature-salience account predicts hair color responses on changed faces in the experimental condition, but not in the control condition. The associative-mediation account predicts no hair color responses in either condition for changed face stimuli. If hair color is the dominant feature regardless of equivalence training, changed face stimuli should generate hair color responses in both conditions.

Table 2 gives the number of participants that gave 0, 1 or 2 ''hair color'' responses for changed faces. In the control condition, participants tended to ignore hair color and instead responded by treating the changed face in the same way as the trained face from which it was derived. This result confirms that hair color was not so salient as to swamp all other features, and is consistent with the predictions of both the feature-salience and associative-mediation accounts. Table 2 shows a very similar pattern of behavior for the experimental condition: Most subjects classified the changed face not by hair color but by the face from which it was derived. In other words, even though these subjects had previously learned to treat faces with the same hair color as equivalent, this single diagnostic feature was not enough to determine how a changed face would be classified. There was no difference between the patterns of responding to changed faces in the experimental and control conditions, $\chi^2(1) = 0.19$, $p > .5$ one-tailed. Taking both conditions together, fewer participants responded based on hair color to the changed faces than based on face identity, $\chi^2(1) = 3.13$, $p = .04$, one-tailed.

Discussion

The results of Experiment 3 go counter to the feature salience account. When the diagnostic feature that determines equivalence classes (i.e., hair color) was changed, the responses to the stimulus followed the equivalence class in most participants, not the new single-feature change. This shows at least that attention paid to the diagnostic feature did not outweigh associations to the stimulus as a whole. A caveat is that in the Fish task, there are two stimuli in each stimulus class. It is possible that feature salience does play a role when larger stimulus classes are used. With larger stimulus classes, shared features and not individual stimuli may become the basis for associations with outcomes.

The results are consistent with the associative mediation account. They are not strong evidence for it, however, as in both conditions a minority of participants responded in line with the hair color feature, and not face identity. This is unsurprising insofar as a trained face with changed hair color is a degraded version of the trained stimulus and, as such, would be expected to be less able to retrieve the consequent associated with the original face. Nevertheless, it is fair to say that the associative mediation account is supported more by the elimination of a rival explanation than by confirmation of its prediction.

GENERAL DISCUSSION

In acquired equivalence, two or more stimuli become functionally equivalent through pairing with the same outcome. Experiments 1 and 2 showed that such equivalence training can alter the way stimuli are stored in memory. The training made equivalent faces more similar and therefore more confusable, as shown by the pattern of errors in an associative recognition test. Experiment 3 showed that representational change is more likely to take the form of mediated associations (Hall et al., 2003) than

of an increased attention to and reliance on diagnostic features.

By showing that acquired equivalence training influences the formation of episodic memories, our findings suggest that acquired equivalence training and episodic memory rely on overlapping memory systems. Gluck and Myers (2001; also see Hodder, Gearge, Killcross, & Honey, 2003; Myers & Gluck, 1996) proposed that such representational modifications might occur in the hippocampal region. Specifically, representations might be changed to increase generalization between stimuli that co-occur or are paired with the same outcome (consequent). Prior work showed that humans with hippocampal atrophy are able to learn the initial face–fish mappings, and are able to continue to respond correctly to trained pairs at test, but are impaired at the novel pairs in a transfer test—failing to generalize and pair A2 with X2 and B2 with Y2 (Myers et al., 2003). Functional neuroimaging (fMRI) studies have also documented hippocampal region activation while healthy adult humans learn a similar task (Preston, Shrager, Dudukovic, & Gabrieli, 2004). Similarly, rats with entorhinal cortex lesions are disrupted at acquired equivalence (Coutureau, Killcross, Good, Marshall, Ward-Robinson, & Honey, 2002), although the same authors found that lesions to the hippocampus proper did not impair acquired equivalence. The fact that our current findings show interactions between equivalence training and episodic memory, usually presumed to rely on the hippocampal region, adds to the body of evidence implicating the hippocampal region in acquired equivalence.

On the other hand, our results from Experiment 3 contrast with the results of Bonardi et al. (2005), who found that equivalence training has much stronger effects when equivalent stimuli share a clear feature than when they do not. Whereas we found little evidence that individual features determined generalization, Bonardi et al. found that a shared feature facilitates equivalence training. However, the fact that such shared features facilitate acquired equivalence does not necessarily imply that acquired equivalence depends on such shared features. It is only the latter idea that was disconfirmed in Experiment 3. Our results are consistent with those of Goyos (2000). He had children identity-match eight stimuli. Correct choices were rewarded with beads of one color for four stimuli, and beads of another color for the other four stimuli. In two experiments, he found that most children who participated would spontaneously name the colors while learning new discriminations involving the stimuli. Those that did not showed no generalization in responding in a transfer test. When taught to name the bead colors during extra training, transfer was normal. This suggests that the verbal labels mediated transfer of function, consistent with the associative-mediation account (although Goyos himself did not interpret his results in such terms).

In summary, the current experiments suggest that equivalence training changes the representations of the stimuli involved. Added associations make equivalent stimuli more similar, and therefore more easily confusable. Moreover, they allow responses learned to one stimulus to generalize to other stimuli.

This view is not very different from how exemplar-based models view performance in categorization tasks (e.g., Nosofsky, 1988; Nosofsky & Palmeri, 1997; Shiffrin, 2003). These models assume that category assignments are stored in the memory representations of stimuli, one of which is stored for each trial. This is equivalent to the creation of an association, assumed by associative mediation, between the stimulus and the outcome or response. Then, at test, presentation of the stimulus evokes retrieval of stimulus representations from memory, and a response is given based on the category assignment stored in the retrieved representations. What would an exemplar model of a standard acquired equivalence task look like? First, stimuli A1 and A2 are coupled to response X1, leading response X1 to be incorporated in, or linked to, memory representations of A1 and A2. Then, A1 is coupled with response X2, and so X2 is incorporated in or linked to the representation of A1. What would now happen if A2 were to be presented, and a choice given between response X2 and a different response, say Z? Retrieving stimulus A2 would also retrieve response X1, which leads to the retrieval of representations of A1, which in turn leads to retrieval of X2. Unless a representation of the alternative response Z is similarly activated, A2 will evoke X2—just as observed in an acquired equivalence study.

Whether or not such an account would be an improvement over a purely associative account of acquired equivalence remains an issue for further study.

REFERENCES

Bonardi, C., Graham, S., & Hall, G. (2005). Acquired distinctiveness and equivalence in human distrimination learning: Evidence for an attentional process. *Psychonomic Bulletin & Review*, *12*, 88–92.

Bonardi, C., Rey, V., Richmond, M., & Hall, G. (1993). Acquired equivalence of cues in pigeon autoshaping: Effects of training with common consequences and common antecedents. *Animal Learning and Behavior*, *21*, 369–376.

Coutureau, E., Killcross, A. S., Good, M., Marshall, V. J., Ward-Robinson, J., & Honey, R. C. (2002). Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 388–396.

Dube, W. V., McIlvane, W. J., Maguire, R. W., Mackay, H. A., & Stoddard, L. T. (1989). Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior*, *51*, 65–76.

Foree, D. D., & LoLordo, V. M. (1970). Attention in the pigeon: Differential effects of food-getting versus shock-avoidance procedures. *Journal of Comparative and Physiological Psychology*, *85*, 551–558.

Gluck, M. A., & Myers, C. E. (2001). *Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus in Learning and Memory*. Cambridge, MA: MIT Press.

Goyos, C. (2000). Equivalence class formation via common reinforcers among preschool children. *Psychological Record*, *50*, 629–654.

Hall, G. C., Mitchell, C., Graham, S., & Lavis, Y. (2003). Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. *Journal of Experimental Psychology: General*, *132*, 266–276.

Hodder, K. I., Gearge, D. N., Killcross, A. S., & Honey, R. C. (2003). Representational blending in human conditional learning: Implications for associative theory. *Quarterly Journal of Experimental Psychology B*, *56*, 223–238.

Honey, R., & Hall, G. (1991). Acquired equivalence and distinctiveness of cues using a sensory-preconditioning procedure. *Quarterly Journal of Experimental Psychology*, *43B*, 121–135.

Honey, R., & Hall, G. C. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 338–346.

Hull, C. L. (1939). The problem of stimulus equivalence in behavior theory. *Psychological Review*, *46*, 9–30.

Kraemer, P. J. (1984). Forgetting of visual discriminations by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 530–542.

Kruschke, J. K. (1979). *Dimensional relevance shifts in category learning*. Indiana University Cognitive Science Program Technical Report #79.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Mackintosh, N. J. (1965). Selective attention in animal discrimination learning. *Psychological Bulletin*, *64*, 124–150.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

Markham, M. R., Dougher, M. J., & Augustson, E. M. (2002). Transfer of operant discrimination and respondent elicitation via emergent relations of compound stimuli. *Psychological Record*, *52*, 325–350.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Myers, C. E., & Gluck, M. (1996). Cortico-hippocampal representations in simultaneous odor discrimination learning: A computational interpretation of Eichenbaum, Mathews & Cohen (1989). *Behavioral Neuroscience*, *110*, 685–706.

Myers, C. E., Shohamy, D., Gluck, M., Grossman, S., Kluger, A., Ferris, S., et al. (2003). Dissociating hippocampal vs. basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience*, *15(2)*, 1–9.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 700–708.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*, 148–152.

Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*, *4*, 203–208.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 803–814.

Schenk, J. J. (1994). Emergent relations of equivalence generated by outcome-specific consequences in conditional discrimination. *Psychological Record*, *44*, 537–558.

Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science*, *27*, 342–378.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*, 23–44.

Tonneau, F. (2001). Equivalence relations: A critical analysis. *European Journal of Behavioral Analysis*, *2*, 1–33.

Urcuioli, P. (2001). Categorization & acquired equivalence. In R. G. Cook (Ed.), *Avian visual cognition [On-line]*. Available: www.pigeon.psy.tufts.edu/avc/urcuioli/.

Urcuioli, P., Lionello-DeNolf, K., Michalek, S., & Vasconcelos, M. (2006). Some tests of response membership in acquired equivalence classes. *Journal of the Experimental Analysis of Behavior*, *86*, 81–107.

Ward-Robinson, J., & Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *Quarterly Journal of Experimental Psychology B, 52*, 335–350.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. New York: McGraw-Hill.

Zeelenberg, R., Boot, I., & Pecher, D. (2005). Activating the critical lure during study is unnecessary for false recognition. *Consciousness and Cognition, 14*, 316–326.

Received: April 18, 2007
Final Acceptance: September 29, 2008